Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2007-29

Bayesian Estimation of the Multi-factor Heston Stochastic Volatility Model

Sylvia Frühwirth-Schnatter and Leopold Sögner[a]

First version October 2007
Final version: November 2008
Published in: Communications in Dependability and Quality Management, *11 (2008), 5–25*

[a]Department of Economics and Finance, Institute for Advanced Studies, A-1060 Vienna, Austria; e-mail address: soegner@ihs.ac.at

Abstract

In this article we investigate a multi-factor version of the Heston (1993) stochastic volatility model. First, we provide explicit expressions for excess kurtosis and autocorrelation of squared returns and show that excess kurtosis is smaller than three and squared autocorrelations are smaller than 0.2 even for a multi-factor model. Then we discuss a fully Bayesian analysis based on Markov chain Monte Carlo (MCMC) estimation and data augmentation and improve the performance of MCMC estimation by using a partially centered parametrization of the model. Finally, we apply the multi-factor Heston stochastic volatility model to simulated as well as to exchange rate data.

Keywords: excess kurtosis, MCMC, parameterization, stochastic volatility, volatility clustering

1 Introduction

Modern option pricing theory often uses models where a closed form expression for the price of a European call option is available. A first yardstick when considering option pricing theory is the Black and Scholes (1973) model, however, this model fails to reproduce important stylized features of financial time series such as excess kurtosis and volatility clustering. To capture such stylized features stochastic volatility models have been introduced such as Gaussian Ornstein-Uhlenbeck processes for the volatility (Stein and Stein, 1991) or for the log volatility (Jacquier, Polson, and Rossi, 1994; Andersen and Lund, 1997; Hull and White, 1987), constant elasticity models (Heston, 1993; Meddahi and Renault, 2000), jump diffusion type models (Duffie, Pan, and Singleton, 2000; Bollerslev and Zhou, 2002), or Ornstein-Uhlenbeck processes driven by a Lévy process (Barndorff-Nielsen and Shephard, 2001).

To account for volatility clustering and excess kurtosis in a more flexible way, several authors (Barndorff-Nielsen and Shephard, 2001; Bai, Russell, and Tiao, 2003; Chernov, Gallant, Ghyseks, and Tauchen, 2003) demand for multi-factor stochastic volatility models because a one factor model often fails to fit periods of high volatility. In this paper we investigate a multi-factor version of the Heston (1993) stochastic volatility model. We derive explicit expressions for excess kurtosis and volatility clustering in the marginal distribution of the returns generated by this model. We show that in contrast to other multi-factor stochastic volatility models the scope for volatility clustering and excess kurtosis is limited for a Heston stochastic volatility model and does not increase when adding additional factors.

As for other stochastic volatility models, see, for instance, Aït-Sahalia (2007) for a recent review, parameter estimation turns out to be a challenge for the multi-factor Heston stochastic volatility. Following the seminal paper by Jacquier et al. (1994) we pursue a Bayesian approach as many other authors did for related models, see e.g. Elerian, Chib, and Shephard (2001), Eraker (2001), Jones (2003), Jacquier, Polson, and Rossi (2004), Roberts, Papaspiliopoulos, and Dellaportas (2004), Griffin and Steel (2006), Bates (2006), Kalogeropoulos (2007), Frühwirth-Schnatter and Sögner (2008), Golightly and Wilkinson (2008), and Strickland, Martin, and Forbes (2008).

Alternative estimation methods that have been or could be applied include simulations (Gallant and Tauchen, 1996), generalized methods of moments (Barndorff-Nielsen and Shephard, 2002; Bollerslev and Zhou, 2002), efficient method of moments (Chernov et al., 2003), spectral GMM (Singleton, 2001; Chacko and Viceira, 2003), non-parametric techniques (Stanton, 1997) and methods based on approximation of the transition densities of the stochastic differential equation (Aït-Sahalia, 2002). In comparison to these approaches Bayesian estimation provides us with a lot of extra information like posterior estimates of the instantaneous volatilities and the increases in integrated volatility.

Bayesian inference for stochastic volatility models is carried out using data augmentation and Markov chain Monte Carlo (MCMC) sampling. Usually, intra-observations times are introduced between each pair of discrete observations and a first order Euler scheme is applied to approximate the underlying transition density. In the present paper we consider for the Heston stochastic volatility model alternative latent processes as missing data, namely the instantaneous and integrated volatilities. When implementing our Metropolis-Hastings algorithm we provide a new solution to the problem that the transition density, in this case the joint density of instantaneous and integrated volatility, is not available in closed form. Since the characteristic function of the transition density is known in closed form, we use a normal approximation of this bivariate density where the first two moments are derived from the characteristic function.

It is well known that the convergence behavior of MCMC sampling crucially depends on the parameterization of the latent process, see, for instance, Papaspiliopoulos, Roberts, and Skold (2003), Roberts et al. (2004), Kalogeropoulos (2007), Frühwirth-Schnatter and Sögner (2008), Golightly and Wilkinson (2008), and Strickland et al. (2008). Also for the Heston model a centered parameterization based on the instantaneous and integrated volatilities leads to a very poor performance of the corresponding MCMC sampling. To make MCMC estimation feasible we consider a partially centered version of the latent process in the spirit of Papaspiliopoulos et al. (2003) and show that the performance of the resulting MCMC sampler improves considerably.

The rest of the paper is organized as follows. In Section 2 we study the multi-factor Heston stochastic volatility model. Section 3 deals with Bayesian estimation using data augmentation and MCMC. Finally in Section 4 we discuss an application to simulated and to exchange rate data.

2 The Multi-factor Heston Stochastic Volatility Model

2.1 Model Formulation

We investigate the following multi-factor stochastic volatility model based on Heston (1993):

$$\mathrm{d}y^{\star}(t) = (\mu + \beta\sigma^2(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_0(t)\ , \tag{1}$$

where the actual volatility $\sigma^2(t)$ is given by a superposition of k independent square root processes:

$$\sigma^2(t) = \sum_{i=1}^{k} \sigma_i^2(t) \ , \tag{2}$$

$$\mathrm{d}\sigma_i^2(t) = \lambda_i(\alpha_i - \sigma_i^2(t))\,\mathrm{d}t + \tau_i\sigma_i(t)\,\mathrm{d}W_i(t), \tag{3}$$

with $W_i(t), i = 0, 1, \ldots, k$ being independent Brownian motions.

We focus on the case, where the parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)'$, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)'$ are unknown and have to be inferred from the data. The parameters μ and β will not be estimated and set to zero. The parameter α_i is the mean of the instantaneous volatility process $\sigma_i^2(t)$ and τ_i is the volatility of volatility parameter. The parameter λ_i is often interpreted as the speed of mean reversion.

If the parameters satisfy $\lambda_i > 0$ and $2\lambda_i\alpha_i/\tau_i^2 \geq 1$ (*Feller condition*), then a stationary law of $\sigma_i^2(t)$ exists, given by the $\mathrm{Gam}(a_i, b_i)$-distribution with $b_i = 2\lambda_i/\tau_i^2$ and $a_i = b_i\alpha_i$, hence $\mathrm{E}(\sigma_i^2(t)) = \alpha_i$, and $\sigma_i^2(t) > 0$ with probability one. The Feller condition implies $\tau_i^2/(2\lambda_i) \leq \alpha_i$ which restricts the volatility of the volatility in the following way: $\mathrm{var}(\sigma_i^2(t)) = \alpha_i\tau_i^2/(2\lambda_i) \leq \alpha_i^2$. We show in the next subsection that this restricts the moments of the distribution of aggregate returns, in particular the kurtosis as well as the autocorrelation of the squared returns in an unexpected way.

2.2 Excess Kurtosis and Volatility Clustering

The need for multi-factor models has been motivated by the fact that one-factor models poorly capture excess kurtosis and volatility clustering (Bai et al., 2003; Chernov et al., 2003). To evaluate the multi-factor Heston stochastic volatility model in this respect we derive explicit expressions for the implied excess kurtosis and volatility clustering of returns observed at regular time points $t_n = n\Delta$. First, we define the integrated volatility:

$$\sigma^{2\star}(t) = \sum_{i=1}^{k} \sigma_i^{2\star}(t), \qquad \sigma_i^{2\star}(t) = \int_0^t \sigma_i^2(u)\,\mathrm{d}u \ ,$$

and the increases in integrated volatility h_n from $t_{n-1} = (n-1)\Delta$ to $t_n = n\Delta$,

$$h_n = \sum_{i=1}^{k} h_{i,n}, \qquad h_{i,n} = \sigma_i^{2\star}(t_n) - \sigma_i^{2\star}(t_{n-1}).$$

One has to distinguish between the conditional and the marginal distribution of the returns y_n. Conditionally on h_n the returns are normally distributed,

$$y_n|h_n \sim \mathcal{N}(0, h_n), \tag{4}$$

while the marginal distribution of y_n where h_n is integrated out is nonnormal. It is possible to derive the moments of this distribution as well as the autocorrelation of y_n and y_n^2 from the moments of h_n, see Appendix B. The first three moments are

given by:

$$\mathrm{E}(y_n) = \mathrm{E}(y_n^3) = 0, \tag{5}$$

$$\mathrm{var}(y_n) = \mathrm{E}(y_n^2) = \mathrm{E}(h_n) = \Delta \sum_{i=1}^{k} \alpha_i. \tag{6}$$

The kurtosis coefficient $\mathrm{kur}(y_n)$ is given by, see (34):

$$\mathrm{kur}(y_n) = 3 + 3\frac{\mathrm{var}(h_n)}{\mathrm{E}(h_n)^2}. \tag{7}$$

The process y_n is uncorrelated, while y_n^2 is correlated, see (35):

$$\mathrm{corr}(y_n^2, y_{n-s}^2) = \frac{\mathrm{cov}(h_n, h_{n-s})}{3\,\mathrm{var}(h_n) + 2\,\mathrm{E}(h_n)^2}. \tag{8}$$

From (7) and (8) it is evident that the Heston model allows for excess kurtosis and volatility clustering, however, the range of these features is rather limited. As shown in Appendix B, equation (27), the coefficient of variation of h_n is bounded:

$$\frac{\mathrm{var}(h_n)}{\mathrm{E}(h_n)^2} \leq 1, \tag{9}$$

which implies an upper bound both for excess kurtosis and volatility clustering. It follows immediately from (7) that the kurtosis coefficient lies in the interval [3,6] and excess kurtosis is limited by 3 even for a multi-factor model. Furthermore, the bound (9) implies $\mathrm{E}(h_n)^2 \geq \mathrm{var}(h_n)$ and therefore:

$$\mathrm{corr}(y_n^2, y_{n-s}^2) \leq \frac{\mathrm{cov}(h_n, h_{n-s})}{3\,\mathrm{var}(h_n) + 2\,\mathrm{var}(h_n)} = \frac{1}{5}\,\mathrm{corr}(h_n, h_{n-s}). \tag{10}$$

Thus the autocorrelation function of the squared returns is limited by 1/5.

In contrast to other multi-factor stochastic volatility models like the model suggested in Barndorff-Nielsen and Shephard (2001), where a multi-factor model increases the flexibility of the marginal distribution of the returns significantly, the ability of the Heston model to capture properties of empirical times series does not increase when using multifactor models, since the upper bounds on the kurtosis coefficient and the autocorrelation of the squared returns are independent of the number k of factors. This result implies that the multi-factor Heston stochastic volatility model is likely to be too restrictive for time series that are typically observed in financial markets. It has been observed earlier in the empirical studies by Bollerslev and Zhou (2002) and Jones (2003) that the Heston model is inferior to other stochastic volatility models such as the CEV model and the two-gamma model.

3 Bayesian Estimation of the Heston Model

Let $\boldsymbol{y} = \{y_n\}_1^N$ denote a sequence of N returns observed at regular intervals within the time span $[0, T]$, where $T = N\Delta$. The same time grid is applied to the instantaneous volatilities $\sigma_i^2(t)$. Thus $\{\sigma_{i,n}^2\}_0^N$ are $N+1$ instantaneous volatilities in the

time span $[0, T]$ observed at equidistant intervals Δ, whereas $\sigma_n^2 = \sum_{i=1}^k \sigma_{i,n}^2$ is the actual volatility, see (2). The whole vector of unknown parameters is abbreviated by $\boldsymbol{\theta}$, whereas $\boldsymbol{\theta}_i = (\alpha_i, \lambda_i, \tau_i)'$ refers to the parameters of factor i. The goal of this section is to describe how the model parameters $\boldsymbol{\theta}$ can be estimated from the observed returns $\boldsymbol{y}$ by a Bayesian approach using data augmentation and MCMC estimation.

3.1 Parameterization of the latent volatility processes

Bayesian estimation of $\boldsymbol{\theta}$ given $\boldsymbol{y}$ is based on data augmentation by introducing a latent processes $\boldsymbol{X}$ which together with $\boldsymbol{\theta}$ constitites our set of unknown quantities that need to be inferred from the data. An MCMC scheme is then constructed based on sampling alternatively from $p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y})$ and $p(\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{y})$.

It is a well known that the appropriate choice of the latent process $\boldsymbol{X}$ may crucially influence the convergence behavior of the resulting MCMC sampler. Papaspiliopoulos et al. (2003) and Roberts et al. (2004) found that a non-centered parameterization based on choosing latents processes $\boldsymbol{X}$ that are independent of the unknown model parameters $\boldsymbol{\theta}$ often exhibits superior sampling properties. Refinements and applications of this concept are provided in Kalogeropoulos (2007), Frühwirth-Schnatter and Sögner (2008), Golightly and Wilkinson (2008), and Strickland et al. (2008).

In the context of the Heston stochastic volatility model natural candidates for the latent processes are the integrated volatilities and the instantaneous volatilities, suggesting to choose $\boldsymbol{X} = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N, \boldsymbol{\sigma}_0^2, \ldots, \boldsymbol{\sigma}_N^2)$, where $\boldsymbol{h}_n = (h_{1,n}, \ldots, h_{k,n})'$ and $\boldsymbol{\sigma}_n^2 = (\sigma_{1,n}^2, \ldots, \sigma_{k,n}^2)'$, respectively. However, as expected from Papaspiliopoulos et al. (2003) such a centered parameterization leads to a very poor performance of the resulting MCMC scheme.

One could define a non-centered parameterization in the spirit of Roberts et al. (2004) by considering following standardized random variables $h_{i,n,NC}$ and $\sigma_{i,n,NC}^2$:

$$h_{i,n,NC} = \frac{h_{i,n} - a_i}{A_i}, \qquad \sigma_{i,n,NC}^2 = \frac{\sigma_{i,n}^2 - b_i}{B_i}, \tag{11}$$

where $a_i = \alpha_i \Delta$ and $b_i = \alpha_i$ are the expected values of $h_{i,n}$ and $\sigma_{i,n}^2$ and $B_i = \mathrm{var}(\sigma_{i,n}^2)^{1/2}$, where $\mathrm{var}(\sigma_{i,n}^2) = \alpha_i \tau_i^2 / (2\lambda_i)$, while $A_i = \Delta B_i$ approximates the standard deviation of $h_{i,n}$. However, also a non-centered parameterization based on choosing $h_{i,n,NC}$ and $\sigma_{i,n,NC}^2$ as latent variables leads to a very poor performance of the resulting MCMC scheme. Also Papaspiliopoulos et al. (2003) found examples of normal hierarchical models where the non-centered parameterization did not lead to good sampling properties. For the normal hierarchical model the convergence rate of either parameterization depends on the proportion of volatility in the observation equation and in the latent equation, respectively. This motivated Papaspiliopoulos et al. (2003) to construct a partially centered parameterization.

In this paper we follow their approach and propose a partially centered parameterization with respect both to the mean and the variance.[1] Partially centering of a

[1] In a prior version of this paper Frühwirth-Schnatter and Sögner (2002) developed another partially centered parameterization with respect to location. In this approach the dimension of

stochastic volatility model is based on subtracting from $\sigma_{i,n}^2$ the fraction $\nu_i b_i$, where $\nu_i \in [0,1]$, and dividing the resulting variable by the fractional power $B_i^{\gamma_i}$, where $\gamma_i \in [0,1]$. Similar considerations apply to the increases in integrated volatility $h_{i,n}$. This leads to following partially centered volatilities:

$$h_{i,n,PC} = \frac{h_{i,n} - \nu_i a_i}{A_i^{\gamma_i}}, \qquad \sigma_{i,n,PC}^2 = \frac{\sigma_{i,n}^2 - \nu_i b_i}{B_i^{\gamma_i}} \ . \tag{12}$$

For $\nu_i = \gamma_i = 0$ this degenerates to the centered parameterization, while the non-centered parameterization (11) results with $\nu_i = \gamma_i = 1$.

A certain drawback of the parameterization (12) is that the domain of $(\sigma_{i,n,PC}^2, h_{i,n,PC})$ is $\Re^2$, while $\sigma_{i,n}^2$ and $h_{i,n}$ are strictly larger than zero. To ensure that for a certain configuration of $(\sigma_{i,n,PC}^2, h_{i,n,PC})$ and $(\alpha_i, \tau_i, \lambda_i)$ the implied values of $\sigma_{i,n}^2$ and $h_{i,n}$ are positive, the MCMC sampler rejects draws for which this is not the case.

Let us relate our approach to the current literature. As regards centering with respect to volatility, Roberts and Stramer (2001) propose a transformation to obtain a constant variance setting for non-linear diffusions. We do not apply this transformation for the following reasons: first, the transformation of Roberts and Stramer (2001) is path dependent, while we obtain another square root process when multiplying the centered square root process with the constant $1/B_i^{\gamma_i}$. Second, it turned out that working with $\gamma_i = 1$ results in MCMC samples with downward biased volatilities of volatilities. Kalogeropoulos (2007) extended the methodology of Roberts and Stramer (2001) to multivariate diffusions and apply this concept also to the Heston model, with the result that the autocorrelations of the MCMC paths decrease, however the state-space equation becomes highly non-linear. In this paper we construct an affine transformation, leading to similar serial correlation in the MCMC samples.

Golightly and Wilkinson (2008) propose a methodology for non-linear diffusions observed with noise to break down the dependence structure between the diffusion coefficient and the quadratic variation. By applying the Euler scheme, the latent process is transformed in such a way that sampling from diffusion bridges becomes possible. Strickland et al. (2008) develop non-centering with respect to mean and scale for a Heston type stochastic volatility model and apply affine transformations different from those used in this paper to the volatility process.

3.2 Prior Distributions

Instead of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, we consider a different parameterization where $\tilde{\boldsymbol{\alpha}} = (\sum_{i=1}^k \alpha_i, \alpha_2, \ldots, \alpha_k)'$. We use the following prior distribution:

$$p(\boldsymbol{\theta}) \propto p(\tilde{\boldsymbol{\alpha}}) \prod_{i=1}^{k} p(\lambda_i|\tau_i) p(\tau_i) \mathbf{1}_{\{2\lambda_i\alpha_i/\tau_i^2 \geq 1\}} \ , \tag{13}$$

where $\tau_i \sim \text{Gam}(d_0, D_0)$-distribution and $\lambda_i|\tau_i \sim \text{Gam}(l_0/\tau_i, L_0)$. We set $d_0 = 0.2$, $D_0 = 0.2$, $l_0 = 0.05$, and $L_0 = 2$ in our applications. The indicator function in (13) accounts for the Feller condition

the parameters space has been enlarged. This is not the case with the approach developed in this article.

For $\alpha_2, \ldots, \alpha_k$ we use independent non-informative truncated normal priors, $\alpha_i \sim \mathcal{N}(1, 1000)\mathbf{1}_{(\alpha_i>0)}$, $i = 2, \ldots, k$. We assume $\sum_{i=1}^k \alpha_i | \alpha_2, \ldots, \alpha_k \sim \mathcal{N}(a_0, A_0)\mathbf{1}_{(\alpha_1>0)}$, where $a_0 = \sum_{n=1}^N y_n^2/(N\Delta)$ is a moment estimator of $\sum_{i=1}^k \alpha_i$ and the prior is comparatively informative with $A_0 = a_0 0.025$. a_0 is obtained by estimating the quadratic variation by $\sum_{n=1}^N y_n^2$, and using the fact that the expected integrated volatility on $[0, N\Delta]$ is equal to $N\Delta \sum_{i=1}^k \alpha_i$.

Finally, for the instantaneous volatility process at $t = 0$, we assume $\sigma_{0,i}^2 \sim \text{Gam}(s_0, S_0)$, where $s_0 = 1/10000$ and $S_0 = 1/1000$.

3.3 MCMC Estimation

In the partially centered parameterization introduced in Subsection 3.1, the latent process equals $\boldsymbol{X} = (\boldsymbol{h}_{1,PC}, \ldots, \boldsymbol{h}_{N,PC}, \boldsymbol{\sigma}_0^2, \boldsymbol{\sigma}_{1,PC}^2, \ldots, \boldsymbol{\sigma}_{N,PC}^2)$ where $\boldsymbol{h}_{n,PC} = (h_{1,n,PC}, \ldots, h_{k,n,PC})'$ and $\boldsymbol{\sigma}_{n,PC}^2 = (\sigma_{1,n,PC}^2, \ldots, \sigma_{k,n,PC}^2)'$. By Bayes theorem, the joint posterior distribution of $\boldsymbol{X}$ and $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{X}, \boldsymbol{\theta} | \boldsymbol{y}) \propto p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\theta}) p(\boldsymbol{X} | \boldsymbol{\theta}) p(\boldsymbol{\theta}),$$

where $p(\boldsymbol{\theta})$ is the prior distribution discussed in Subsection 3.2.

The complete data likelihood $p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\theta})$ is easily obtained from (4) as the product of N densities from a normal distribution, where by the definition of the partially centered parameterization $h_{i,n} = A_i^{\gamma_i} h_{i,n,PC} + \nu_i a_i$.

The density $p(\boldsymbol{X} | \boldsymbol{\theta})$ is given by

$$p(\boldsymbol{X} | \boldsymbol{\theta}) = \prod_{i=1}^k \left(p(\sigma_{i,0}^2 | \boldsymbol{\theta}_i) \prod_{n=1}^N p(h_{i,n,PC}, \sigma_{i,n,PC}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i) \right) , \tag{14}$$

where $\sigma_{i,n}^2 = B_i^{\gamma_i} \sigma_{i,n,PC}^2 + \nu_i b_i$. The choice of the prior distributions $p(\sigma_{i,0}^2 | \boldsymbol{\theta}_i)$ has been discussed in Subsection 3.2. The transition density of the partially centered volatilities is related to the transition density of the original stochastic volatility model by the density transformation formula:

$$p(h_{i,n,PC}, \sigma_{i,n,PC}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i) \;=\; p(h_{i,n}, \sigma_{i,n}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)(A_i B_i)^{\gamma_i} \; .$$

The transition density $p(h_{i,n}, \sigma_{i,n}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)$ cannot be derived analytically, but the Fourier transform $\phi(u, v | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)$ is available in closed form, see Appendix A. The inversion formula (see e.g. Klenke (2008))

$$p(h_{i,n}, \sigma_{i,n}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i) \;=\; \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\imath(u h_{i,n} + v\sigma_{i,n}^2)} \phi(u, v | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i) \; \mathrm{d}u \, \mathrm{d}v$$

could be applied to compute the functional value $p(h_{i,n}, \sigma_{i,n}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)$, however, since integration cannot be performed analytically, bivariate numerical integration has to be applied for each $i = 1, \ldots, k$ and $n = 1, \ldots, N$, to obtain $p(\boldsymbol{X} | \boldsymbol{\theta})$. This renders the application of a Metropolis Hastings algorithm to sample from $p(\boldsymbol{X}, \boldsymbol{\theta} | \boldsymbol{y})$ extremely slow. Therefore, we approximate each transition density $p(h_{i,n}, \sigma_{i,n}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)$ by a bivariate normal density, where we use the characteristic function $\phi(u, v | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)$ to match the first two moments, see Appendix D for more details.

To sample $\boldsymbol{X}$ and $\boldsymbol{\theta}$ from the posterior distribution $p(\boldsymbol{X}, \boldsymbol{\theta}|\boldsymbol{y})$, a Markov chain Monte Carlo sampler is constructed which samples from the following conditional densities using a Metropolis Hastings (MH) algorithm:

Step 1: Sample $\tilde{\boldsymbol{\alpha}}$ from $p(\tilde{\boldsymbol{\alpha}}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\lambda}, \boldsymbol{\tau})$.

Step 2: For $i = 1, \ldots, k$, sample (λ_i, τ_i) from $p(\lambda_i, \tau_i|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\alpha}, \boldsymbol{\lambda}_{-i}, \boldsymbol{\tau}_{-i})$.

Step 3: Sample $\boldsymbol{X}$ from $p(\boldsymbol{X}|\boldsymbol{y}, \boldsymbol{\theta})$.

In the first step, we update $\tilde{\boldsymbol{\alpha}} = (\sum_i \alpha_i, \alpha_2, \ldots, \alpha_k)'$ in one block, using the following random walk proposals. A new value for $\sum_i \alpha_i$ is proposed from $(\sum_i \alpha_i^{old}) \exp(c_\alpha \zeta_1)$, whereas a new value for α_i, $i = 2, \ldots, k$ is proposed from $\alpha_i^{old} \exp(c_\alpha \zeta_i)$, where $\zeta_i \sim \mathcal{N}(0, 1)$ for $i = 1, \ldots, k$ and $c_\alpha = 0.1$. In the second step, λ_i and τ_i are updated jointly for each factor i by proposing new values for λ_i and τ_i from $\lambda_i^{old} \exp(c_\lambda \varepsilon_{\lambda,i})$ and $\tau_i^{old} \exp(c_\tau \varepsilon_{\tau,i})$, respectively, where $\varepsilon_{\lambda,i}$ and $\varepsilon_{\tau,i}$ are i.i.d. $\mathcal{N}(0, 1)$ and $c_\lambda = c_\tau = 0.1$.

Updating of the latent process $\boldsymbol{X}$ in the third step is performed on a factor wise basis for each $i = 1, \ldots, k$ by updating all volatilities $(h_{i,n,PC}, \sigma^2_{i,n,PC})$ on a whole block $n_1 \leq n \leq n_1 + B - 1$ of size B. Each pair $(h_{i,n,PC}, \sigma^2_{i,n,PC})$ is proposed from $\mathcal{N}_2((h^{old}_{i,n,PC}, \sigma^{2,old}_{i,n,PC})', \boldsymbol{C}_i)$ with $\boldsymbol{C}_i = 0.1 \operatorname{diag}(\max\{\bar{h}^{old}_{i,n,PC}, 0.5 \sqrt{\operatorname{var}(h^{old}_{i,n,PC})}\}, \max\{\bar{\sigma}^{2,old}_{i,n,PC}, 0.5\sqrt{\operatorname{var}(\sigma^{2,old}_{i,n,PC})}\})$, where $\bar{h}^{old}_{i,n,PC}$ and $\operatorname{var}(h^{old}_{i,n,PC})$ are sample mean and the sample variance of $h^{old}_{i,n,PC}$ over $n = 1, \ldots, N$, and similarly for $\bar{\sigma}^{2,old}_{i,n,PC}$ and $\operatorname{var}(\sigma^{2,old}_{i,n,PC})$. The starting value $\sigma^2_{i,0}$ is updated whenever $n_1 = 1$.

Concerning the blocking strategy, we choose with equal probability between a complete update of the latent process and an update of randomly selected number of blocks. In the former case we draw the block sizes from a Poisson distribution with expectation 10 until the whole path has been partitioned. For the latter updating scheme we draw the number of blocks from a Poisson distribution with expectation five. For each block n_1 is derived from a uniform distribution on the grid $1, \ldots, N-1$ and the block size is once again Poisson with expectation 10.

4 Results

4.1 Performance in simulated data and the level of centering

First we test and tune our sampler for simulated data with $N = 500$ observations generated from a one factor model using two sets of parameters given in Table 1. For the centered and the non-centered parameterization MCMC convergence properties are extremely poor which motivated us to construct the partially centered parameterization introduced in Subsection 3.1. The application of this parameterization requires tuning of ν and γ.

Roberts and Stramer (2001) and Papaspiliopoulos et al. (2003) derive tools to calculate an at least approximately optimal level of partially centering with respect to the mean for hierarchical models. For discretely observed diffusions with non-constant "volatility of volatility" finding such a transformation is not evident, see e.g. the discussion of Wilkinson (2003). In Appendix C we extend the methodology of Papaspiliopoulos et al. (2003) to stochastic volatility models to derive an optimal degree of centering with respect to the mean for a given level γ of centering with

respect to the variance and obtain following result:

$$\nu \;=\; \frac{2(\tilde{a}-1)(\tilde{a}-2)}{2(\tilde{a}-1)(\tilde{a}-2)+\tilde{a}}\;, \tag{15}$$

if $\tilde{a} = 2\alpha\lambda/\tau^2 > 2$, and $\nu = 0$ otherwise. Interestingly, optimal centering with respect to the mean is independent of γ. Note that the condition for partially non-centering is stronger than the Feller condition which implies $2\alpha\lambda/\tau^2 \geq 1$.

ν is monotonically increasing in $\tilde{a}$ and the partial derivatives are positive with respect to α and λ and negative with respect to τ. That means the smaller the volatility of volatility τ and the higher the level α the more non-centering with respect to the mean is required. Furthermore the speed of mean reversion $\exp(-\lambda\Delta)$ has an influence on how much centering with respect to the mean is required. The higher the persistence, i.e. the smaller λ, the more centering is required. Similar results were obtained by Pitt and Shephard (1999) and Frühwirth-Schnatter (2004) for discrete time, time-varying parameter models.

The optimal level of centering with respect to the variance, controlled by γ, is derived by means of tuning and comparing the efficiency of the resulting samplers. For the setting where $\lambda = 0.2$, $\tau = 0.2$ and $\alpha = 0.35$ rule (15) recommends approximately $\nu = 0.8$. Tuning of γ led to a value of $\gamma = 0.35$. For the larger speed of mean reversion setting where $\alpha = 0.35$, $\lambda = 1.5$ and $\tau = 0.5$ rule (15) recommends approximately $\nu = 0.9$. By tuning we found that γ in the range of 0.75 is necessary to obtain reasonable sampling properties.

MCMC sampling was carried out for 50,000 steps after a burn-in of 30,000 steps. The MH acceptance rates for $\boldsymbol{X}$ are approximately 7%, while we observed acceptance rates between 30% and 40% for the components of $\boldsymbol{\theta}$. Although the reparameterization improves the sampling properties compared to completely centered or non-centered parameterizations, the autocorrelations of the simulated paths still remain high. The first order autocorrelations of α lie in the range of 0.6 to 0.7. For λ and τ first order autocorrelations lie in the range of 0.8 to 0.85 for the low speed of mean reversion setting ($\lambda = 0.2$) and in the range of 0.97 to 0.975 for the high speed of mean reversion the setting ($\lambda = 1.5$).

The resulting parameter estimation is reported in Table 1. The estimators based on the posterior mean are near the true values and the standard deviations derived from the variance-covariance matrix of the MCMC draws indicate that the estimators are quite precise. The estimator for λ is more efficient for the setting with low volatility and low speed of mean reversion ($\lambda = 0.2$, $\tau = 0.2$) than for the high speed of mean reversion/high volatility case ($\lambda = 1.5$, $\tau = 0.5$).

Moreover we analyze the simulated paths of σ_n^2. For each path we calculate the mean $\bar{\sigma}_n^2$ and the variance $\widehat{\text{var}}(\sigma_n^2)$. We observe that the average of $\bar{\sigma}_n^2$, $ave(\bar{\sigma}_n^2)$, is close to the true stationary mean $E(\sigma_n^2) = \alpha$. A comparison of the average of $\widehat{\text{var}}(\sigma_n^2)$, $ave(\widehat{\text{var}}(\sigma_n^2))$, and the true stationary volatility $\text{var}(\sigma_n^2) = \alpha\tau^2/(2\lambda)$ indicates that the simulated latent paths exhibit a slightly too small volatility. By a careful look we observe a trade-off between volatility of volatility τ and the speed of mean reversion λ, with $\hat{\tau}$ being too small and $\hat{\lambda}$ being too large. Here the sampler tends to "mix-up" mean reversion with noise.

Parameters				Latent Process	
α	λ	τ	$\frac{\alpha\tau^2}{2\lambda}$	$E(\sigma_n^2)$	$\text{var}(\sigma_n^2)$
0.35	0.2	0.2	0.035	0.35	0.035
$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\tau}$	$ave(\frac{\alpha\tau^2}{2\lambda})$	$ave(\bar{\sigma}_n^2)$	$ave(\widehat{\text{var}}(\sigma_n^2))$
0.3551	0.2396	0.1746	0.0228	0.3611	0.0188
(0.0738)	(0.0268)	(0.0090)	(0.0026)	(0.0231)	(0.0017)
α	λ	τ	$\frac{\alpha\tau^2}{2\lambda}$	$E(\sigma_n^2)$	$\text{var}(\sigma_n^2)$
0.35	1.5	0.5	0.0292	0.35	0.0292
$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\tau}$	$ave(\frac{\alpha\tau^2}{2\lambda})$	$ave(\bar{\sigma}_n^2)$	$ave(\widehat{\text{var}}(\sigma_n^2))$
0.3534	1.6058	0.4233	0.0198	0.3420	0.0190
(0.0035)	(0.1440)	(0.0271)	(0.0026)	(0.0142)	(0.0039)

Table 1: Performance in simulated data for two parameter settings. Posterior means and standard deviations (in parentheses) are based on 50,000 MCMC draws (after 30,000 burn-in steps) using a partially centered parameterization with $\nu = 0.8$ and $\gamma = 0.35$ for $\lambda = 0.2$ and $\tau = 0.2$ and $\nu = 0.9$ and $\gamma = 0.75$ for $\lambda = 1.5$ and $\tau = 0.5$.

4.2 Application to Daily DM/US$ Exchange Rate Data

We apply our estimation methodology to daily DM/US$ exchange rate data from January 1997 to December 1998 where $N = 500$.[2] Squared returns are shown in the left hand panel of Figure 1.

The main reason behind choosing this particular data set is the limitation of the Heston model to explain higher levels of excess kurtosis and volatility clustering as discussed in Subsection 2.2. Both the level of volatility clustering and excess kurtosis is relatively low for this data set compared to time series data usually observed on financial markets with the lag 1 autocorrelation of the squared returns equals 0.0576 and the kurtosis coefficient equals 3.5.

4.2.1 Bayesian Parameter estimation

MCMC estimation of a one and a two factor model was carried out for 500,000 steps, where we discarded the first 200,000 samples as burn-in. Estimation is based on a partially centered version where $\gamma = 0.35$ and $\nu = 0.8$ for the one factor model, while $(\gamma_1, \gamma_2)' = (0.35, 0.75)$ and $(\nu_1, \nu_2)' = (0.9, 0.9)'$ for the two factor setting.

Table 2 presents the means and the standard deviations of the marginal posteriors of the model parameters for the one factor and the two factor setting. For the two factor model we obtain $\lambda_1 < \lambda_2$, i.e. the persistence of the first factor is larger than that of the second one, $\exp(-\hat{\lambda}_1) = 0.92$ versus $\exp(-\hat{\lambda}_2) = 0.35$. The first factor describes the persistent part of the exchange rate volatility, while the second factor is responsible for innovations of lower memory.

In Figure 1 we present the posterior mean of the instantaneous volatility process for the two factor setting. The dash lines in Figure 1 represent a confidence interval, obtained by taking the posterior mean $\pm$ posterior standard deviation of the

[2]The data were downloaded from the Deutsche Bundesbank
http://www.bundesbank.de/statistik/statistik_zeitreihen.php?func=row&tr=wt5009

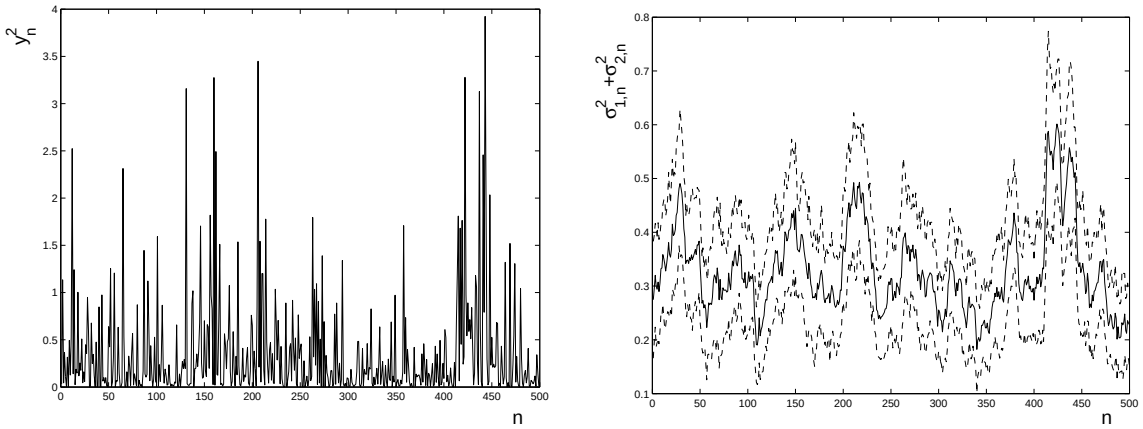

Figure 1: Daily squared returns y_n^2 of DM/US$ exchange rate data from January 1997 to December 1998 (left hand panel); posterior means (solid line) and confidence intervals (means $\pm$ standard deviations; dashed lines) of the instantaneous volatility samples for a two factor Heston stochastic volatility model obtained from $500,000$ simulation steps after $200,000$ burn-in steps (right hand panel).

instantaneous volatility process. When comparing both figures we observe that the estimated instantaneous volatility path and the squared returns have a "similar" pattern. However, the volatility path is much smoother, which is, of course, caused by the autoregressive structure of the volatility process.

4.2.2 Model Validation

Table 3 displays several descriptive statistics of the observed time series. To evaluate whether the one and the two factor Heston stochastic volatility model is able to reproduce these features, we derive the posterior distribution of these statistics using Bayesian predictive methods as in Gelman, Meng, and Stern (1996). Based on the MCMC samples from the posterior density $p(\boldsymbol{\theta}, \boldsymbol{X}|\boldsymbol{y})$ of a one or a two factor model, we simulate for each MCMC draw $\boldsymbol{\theta}$ a return time series with $N = 500$ observations and compute the corresponding statistics. This Monte Carlo experiment results in a distribution of the skewness coefficient $\text{skew}(y_n)$, the kurtosis coefficient $\text{kur}(y_n)$, the autocorrelation function $\text{corr}(y_n, y_{n-s})$ of the returns and the autocorrelation function $\text{corr}(y_n^2, y_{n-s}^2)$ of the squared returns. A model is able to reproduce the characteristics of the empirical time series, if the descriptive statistics of the observed time series are covered by these distributions.

Table 3 shows for several statistics that the observed values are well covered by these distributions both for a one and a two factor model. In addition we calculated autocorrelations of higher order. Here, the decay in autocorrelations implied by the model is stronger than the decay in the empirical autocorrelations of the squared returns and the observed values were not covered by the simulated distributions.

As expected from Subsection 2.2, a two factor model does not improve the fit regarding the marginal distribution of the exchange rate returns. However, this does not mean that the second factor is not necessary. The MCMC sampler has identified significantly different estimates of the speed of mean reversion λ_1 and λ_2 for the two

$k=1$					
$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\tau}$			
0.3542	0.0827	0.0680			
(0.0092)	(0.0205)	(0.0103)			
$k=2$					
$\hat{\alpha}_1$	$\hat{\lambda}_1$	$\hat{\tau}_1$	$\hat{\alpha}_2$	$\hat{\lambda}_2$	$\hat{\tau}_2$
0.1931	0.0811	0.1031	0.1571	1.0402	0.1308
(0.0236)	(0.0266)	(0.0107)	(0.0236)	(0.2104)	(0.0468)

Table 2: Daily DM/US$ exchange rates from January 1998 to December 1998. Posterior means and standard deviations (in parentheses) are based on 500,000 MCMC draws (after 200,000 burn-in steps) using a partially centered parameterization with $\nu = 0.8$ and $\gamma = 0.35$ for a one factor model and $\nu = (0.9, 0.9)'$ and $\gamma = (0.35, 0.75)'$ for a two factor model.

	skew(y_n)	kur(y_n)	corr(y_n, y_{n-1})	corr(y_n^2, y_{n-1}^2)
empirical data	-0.2120	3.4999	-0.0201	0.0576
$k=1$	-7.2e-5	3.1921	-0.0020	0.0257
	(0.1243)	(0.2946)	(0.0458)	(0.0500)
$k=2$	0.0019	3.1906	-0.0018	0.0196
	(0.1256)	(0.2866)	(0.0459)	(0.0483)

Table 3: Time series properties of empirical DM/US$ exchange rate data and simulated statistics (mean values, standard deviations in parentheses) from the posterior of a one factor and a two factor setting.

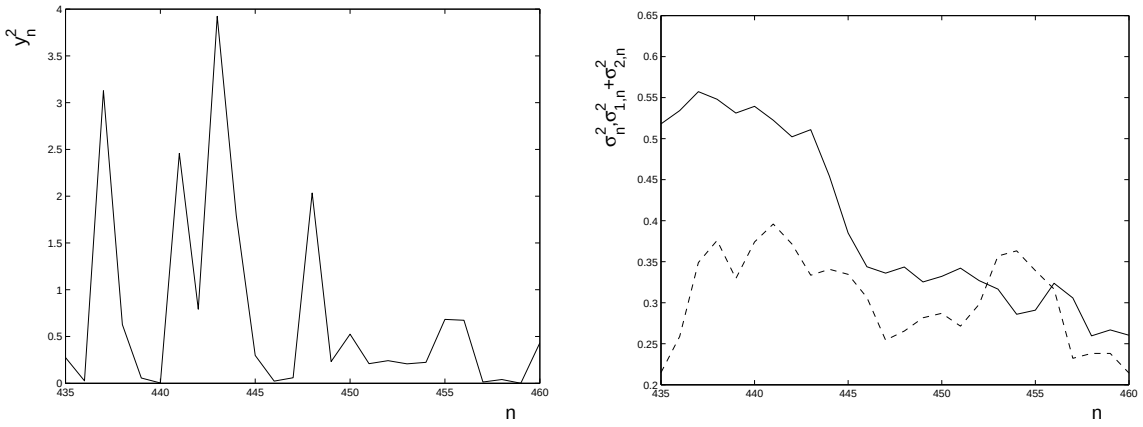

Figure 2: Squared returns y_n^2 of daily DM/US$ exchange for the period $n = 435$ to $n = 460$ (left hand panel). Posterior means of the instantaneous volatility samples for a one factor model (dashed line) and a two factor model (solid line) derived from $500,000$ simulation steps after $200,000$ burn-in steps (right hand panel).

factor setting, see again Table 2. Therefore, we expect that the sampler has inferred additional information, which is not represented by the analysis in Table 3.

The point is that extreme realizations of the squared returns y_n^2 have a relatively low persistence for these data. A one factor model cannot account for this behavior. Figure 2 illustrates this point for the subperiod from $n = 435$ to $n = 460$. We observe spikes in the squared returns for $n = 437$, 441, 443 and 448. Between these observations and after observation $n = 448$, we observe moderate squared returns. The dashed and the solid line in the right panel of Figure 2 plot the sample mean of the instantaneous volatility for a one and a two factor model. For a one factor model σ_n^2 reacts with short delay to larger observations of y_n^2 due to the persistence implied by λ. After extreme yields, the estimates of σ_n^2 return to a level around the mean $\alpha = 0.35$. For the two factor model one factor accounts for changes in volatility, while the second accounts for the baseline volatility. Therefore, the estimates of $\sigma_{1,n}^2 + \sigma_{2,n}^2$ increase with the observations $n = 437$, 441, 443 and 448. The increases in σ_n^2 are less pronounced with the observations $n = 443$ and 448. Here the estimated levels of volatility in the last period, σ_{n-1}^2, are already high enough to cope with the next realization of y_n^2. Summing up, although the advantage of a two factor model cannot be observed with the marginal distribution of the returns, a more plausible behavior after extreme events can be derived with a two factor setting.

5 Concluding Remarks

This article has implemented a Bayesian analysis of the Heston (1993) stochastic volatility model using data augmentation and MCMC. We observed that different parameterizations of the latent volatility process and the parameters of the volatility process result in very different convergence behavior of the MCMC sampler. A standard MCMC implementation based on introducing the instantaneous volatilities and

the increases in integrated volatility as latent variables failed completely. MCMC became feasible by considering a partially centered version of the latent process.

To implement the Metropolis-Hastings algorithm within our MCMC scheme we had to come up with an approximation to the joint transition density of the integrated and actual volatility which is not available in closed form. To this aim we used a normal approximation of this bivariate density derived from the characteristic function of the transition density. A direct inversion of the Fourier transform turned out to be inefficient from a computational point of view.

For a larger step-width the methodology of Elerian et al. (2001) or Eraker (2001) can be included into our estimation methodology. Since for a lot of stochastic volatility models only the characteristic function of the latent volatility process is available, e.g. for models based on Barndorff-Nielsen and Shephard (2002), this approach should be useful for Bayesian estimation of a much broader class of stochastic volatility models.

Finally, we proved that the applicability of the Heston stochastic volatility model is limited for typical time series from financial markets and should be applied only to returns with an excess kurtosis smaller than 3 and autocorrelations of squared returns smaller than 0.2. This provides a theoretical foundation for empirical results obtained by Bollerslev and Zhou (2002) and Jones (2003).

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica 70*, 223–264.

Aït-Sahalia, Y. (2007). Estimating continuous-time models using discretely sampled data. In R. Blundell, T. Persson, and W. K. Newey (Eds.), *Advances in Economics and Econometrics, Theory and Applications*, Chapter 9. Cambridge: Cambridge University Press.

Andersen, T. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics 77*, 343–377.

Bai, X., J. R. Russell, and G. C. Tiao (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics 114*, 349–360.

Barndorff-Nielsen, O. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B 63*, 167–241.

Barndorff-Nielsen, O. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B 64*, 253–280.

Bates, D. (2006). Maximum likelihood estimation of latent affine processes. *Review of Financial Studies 19*, 909–965.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 637–659.

Bollerslev, T. and H. Zhou (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics 109*, 33–65.

Chacko, G. and L. M. Viceira (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics 116*, 259–292.

Chernov, M., A. R. Gallant, E. Ghyseks, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics 116*, 225–257.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*, 1343–1376.

Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica 69*, 959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics 19*, 177–191.

Frühwirth-Schnatter, S. (2004). Efficient Bayesian parameter estimation for state space models based on reparameterizations. In A. Harvey, S. J. Koopman, and N. Shephard (Eds.), *State Space and Unobserved Component Models: Theory and Applications*, pp. 123–151. Cambridge: Cambridge University Press.

Frühwirth-Schnatter, S. and L. Sögner (2002). Bayesian estimation of the Heston stochastic volatility model. In U. Leopold-Wildburger, F. Rendl, and G. Wäscher (Eds.), *Operations Research Proceedings 2002 - Selected Papers of the Annual International Conference of the German Operations Research Society*, Wien, pp. 480–485.

Frühwirth-Schnatter, S. and L. Sögner (2008). Bayesian estimation of OU processes with marginal Gamma law. *The Annals of the Institute of Statistical Mathematics*. forthcoming.

Gallant, A. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*, 657–681.

Gelman, A., X.-L. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica 6*, 733–807.

Golightly, A. and D. J. Wilkinson (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics and Data Analysis 52*, 1674–1693.

Griffin, J. E. and M. F. J. Steel (2006). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *Journal of Econometrics 134*, 605–644.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*, 327–343.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatility. *Journal of Finance 42*, 281–300.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 12*, 371–417.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*, 185–212.

Jones, C. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics 116*, 181–224.

Kalogeropoulos, K. (2007). Likelihood-based inference for a class of multivariate diffusions with unobserved paths. *Journal of Statistical Planning and Inference 137*, 3092–3102.

Klenke, A. (2008). *Probability Theory - A Comprehensive Course*. New York: Springer.

Lamberton, D. and B. Lapeyre (1996). *Introduction to Stochastic Calculus – Applied to Finance* (1^{st} ed.). London: Chapman & Hall.

Lamoureux, C. G. and A. Paseka (2005). Information in options and underlying asset dynamics. *Working paper, University of Arizona*.

Meddahi, N. and E. Renault (2000). Temporal aggregation of volatility models. *Working paper CIRANO, Montreal*.

Papaspiliopoulos, O., G. Roberts, and M. Skold (2003). Non-centered parameterizations for hierarchical models and data augmentation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 7*, pp. 307–326. Oxford: Oxford University Press.

Pitt, M. and N. Shephard (1999). Analytic convergence rates and parametrization issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis 20*, 63–85.

Roberts, G. and O. Stramer (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika 88*, 603–621.

Roberts, G. O., O. Papaspiliopoulos, and P. Dellaportas (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society Series B 66*, 369–393.

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics 102*, 111–141.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance 52*, 1973–2002.

Stein, E. and J. Stein (1991). Stock price distributions and stochastic volatility: An analytic approach. *Review of Financial Studies 4*, 727–752.

Strickland, C. M., G. M. Martin, and C. S. Forbes (2008). Parameterisation and efficient MCMC estimation of non-Gaussian state space models. *Computational Statistics and Data Analysis 52*, 2911–2930.

Wilkinson, D. (2003). Discussion of non-centered parameterizations for hierarchical models and data augmentation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 7*, pp. 323–324. Oxford: Oxford University Press.

A Conditional Densities and Moments

For $(h_{i,n}, \sigma^2_{i,n})'$ the characteristic function $\phi(u, v|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i)$ is given by (see Lamberton and Lapeyre (1996, p. 130))[3]:

$$\phi(u, v|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) = \mathrm{E}[\exp(\imath u h_{i,n} + \imath v \sigma^2_{i,n})] \exp(B_{\phi_1}) \exp(-B_{\phi_2}\sigma^2_{i,n-1}) \ ,$$

where $\imath = \sqrt{-1}$ and

$$\begin{aligned} B_{\phi_1} &= \frac{2\alpha_i\lambda_i}{\tau_i^2} \log\left(\frac{2\gamma_i \exp(0.5\Delta(\gamma_i+\lambda_i))}{\tau_i^2(-\imath v)(\exp(\gamma_i\Delta)-1)+\gamma_i-\lambda_i+\exp(\gamma_i\Delta)(\gamma_i+\lambda_i)}\right) \ , \\ B_{\phi_2} &= \frac{(-\imath v)(\gamma_i+\lambda_i+\exp(\gamma_i\Delta)(\gamma_i-\lambda_i))+2(-\imath u)(\exp(\gamma_i\Delta)-1)}{\tau_i^2(-\imath v)(\exp(\gamma_i\Delta)-1)+\gamma_i-\lambda_i+\exp(\gamma_i\Delta)(\gamma_i+\lambda_i)} \ , \\ \gamma_i &= \sqrt{\lambda_i^2 + 2\tau_i^2(-\imath u)} \ . \end{aligned}$$

The conditional expectation of $h_{i,n}, \sigma^2_{i,n}|\sigma^2_{i,n-1}$ is derived by calculating the gradient vector of the characteristic function at $u = v = 0$. This results in

$$\mathrm{E}(h_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) = \alpha_i\Delta + (\sigma^2_{i,n-1} - \alpha_i)\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i} \ , \tag{16}$$

$$\mathrm{E}(\sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) = (\sigma^2_{i,n-1} - \alpha_i)\exp(-\lambda_i\Delta) + \alpha_i \ . \tag{17}$$

The conditional covariance of $h_{i,n}$ and $\sigma^2_{i,n}$ can be derived[4] from the property of characteristic functions that

$$\mathrm{E}(h_{i,n}\sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) = (1/\imath^2)\frac{\partial^2}{\partial u \partial v}\phi(0, 0|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) \ . \tag{18}$$

[3]This textbook provides the Laplace transform, a substitution of the convolution parameters (u, v) by $(-\imath u, -\imath v)$ provides us with the Fourier transform.

[4]Here we used Mathematica 4.2.

Using (18) and the conditional first moments given by (16) results in:

$$\begin{aligned}
&\operatorname{var}(h_{i,n}|\sigma^2_{i,n-1},\boldsymbol{\theta}_i) \qquad (19)\\
&= \frac{\exp(-2\lambda_i\Delta)\tau_i^2}{2\lambda_i^3}\alpha_i[1+4\exp(\lambda_i\Delta)(1+\lambda_i\Delta)+\exp(2\lambda_i\Delta)(-5+2\lambda_i\Delta)]\\
&+\frac{\exp(-2\lambda_i\Delta)\tau_i^2}{2\lambda_i^3}2\sigma^2_{i,n-1}[-1+\exp(2\lambda_i\Delta)-2\exp(\lambda_i\Delta)\lambda_i\Delta],\\
&\operatorname{var}(\sigma^2_{i,n}|\sigma^2_{i,n-1},\boldsymbol{\theta}_i) \qquad (20)\\
&= \frac{\exp(-2\lambda_i\Delta)(-1+\exp(\lambda_i\Delta))\tau_i^2}{2\lambda_i}\left(\alpha_i(-1+\exp(\lambda_i\Delta))+2\sigma^2_{i,n-1}\right),\\
&\operatorname{cov}(h_{i,n},\sigma^2_{i,n}|\sigma^2_{i,n-1},\boldsymbol{\theta}_i) = \frac{\tau_i^2\exp(-2\lambda_i\Delta)}{2\lambda_i^2}\alpha_i(-1+\exp(2\lambda_i\Delta) \qquad (21)\\
&-2\lambda_i\Delta\exp(\lambda_i\Delta))+\frac{\tau_i^2\exp(-2\lambda_i\Delta)}{2\lambda_i^2}2\sigma^2_{i,n-1}(1+\exp(\lambda_i\Delta)(-1+\lambda_i\Delta)).
\end{aligned}$$

Equation (19) exactly corresponds to the conditional variance obtained in Bollerslev and Zhou (2002, p. 57, Eq. A.5).

B Marginal Moments and Autocorrelations

In Appendix B.1 we compute the mean and the variance of the increases in integrated volatility h_n as well as the autocorrelation function of h_n. These results are used in Appendix B.2 to compute the marginal moments of the returns y_n as well as the autocorrelation function of y_n and y_n^2.

B.1 Marginal Moments and Autocovariance of Increases in Integrated Volatility

The marginal moments $\mathrm{E}(h_n)$ and $\operatorname{var}(h_n)$ are obtained from the conditional moments in Appendix A.

B.1.1 Mean and Variance

From (16) and $\mathrm{E}(\sigma^2_{i,n-1}) = \alpha_i$ we obtain:

$$\begin{aligned}
\mathrm{E}(h_n) &= \sum_{i=1}^k \mathrm{E}(\mathrm{E}(h_{i,n}|\sigma^2_{i,n-1})) \qquad (22)\\
&= \Delta\sum_{i=1}^k\alpha_i + \sum_{i=1}^k\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i}\mathrm{E}(\sigma^2_{i,n-1}-\alpha_i) = \Delta\sum_{i=1}^k\alpha_i.
\end{aligned}$$

Since $h_{1,n},\ldots,h_{k,n}$ are independent we obtain

$$\operatorname{var}(h_n) = \sum_{i=1}^k\operatorname{var}(h_{i,n}) = \sum_{i=1}^k\mathrm{E}(\operatorname{var}(h_{i,n}|\sigma^2_{i,n-1})) + \sum_{i=1}^k\operatorname{var}(\mathrm{E}(h_{i,n}|\sigma^2_{i,n-1})).$$

From (16) and $\mathrm{var}(\sigma^2_{i,n-1}) = \alpha_i\tau_i^2/(2\lambda_i)$ we obtain:

$$\mathrm{var}(\mathrm{E}(h_{i,n}|\sigma^2_{i,n-1})) \;=\; \left[\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i}\right]^2 \frac{\alpha_i\tau_i^2}{2\lambda_i}. \tag{23}$$

From (19) we obtain:

$$\mathrm{E}(\mathrm{var}(h_{i,n}|\sigma^2_{i,n-1})) = \frac{\alpha_i\tau_i^2}{\lambda_i^2}\left(\Delta + \frac{4\exp(-\lambda_i\Delta) - \exp(-2\lambda_i\Delta) - 3}{2\lambda_i}\right). \tag{24}$$

(23) and (24) yield:

$$\mathrm{var}(h_{i,n}) = \frac{\alpha_i\tau_i^2}{\lambda_i^3}\left(\exp(-\lambda_i\Delta) - (1-\lambda_i\Delta)\right), \tag{25}$$

therefore:

$$\mathrm{var}(h_n) = \Delta^2 \sum_{i=1}^{k} \alpha_i \frac{\tau_i^2}{2\lambda_i}\frac{2}{\Delta^2\lambda_i^2}\left(\exp(-\lambda_i\Delta) - (1-\lambda_i\Delta)\right).$$

From the series expansion for $\exp(-\lambda_i\Delta)$ and the Leibnitz criterion we obtain $\exp(-\lambda_i\Delta) - (1-\lambda_i\Delta) \le \frac{\lambda_i^2\Delta^2}{2}$. Together with the Feller condition $\tau_i^2/(2\lambda_i) \le \alpha_i$ this yields the following upper bound for $\mathrm{var}(h_n)$:

$$\mathrm{var}(h_n) \le \Delta^2 \sum_{i=1}^{k} \alpha_i^2. \tag{26}$$

The upper bound of $\mathrm{var}(h_n)$ is a lower bound for $\mathrm{E}(h_n)^2$, since

$$\mathrm{E}(h_n)^2 \;=\; \Delta^2(\sum_{i=1}^{k}\alpha_i)^2 \ge \Delta^2 \sum_{i=1}^{k}\alpha_i^2.$$

Therefore the coefficient of variation of h_n is bounded:

$$0 \le \frac{\mathrm{var}(h_n)}{\mathrm{E}(h_n)^2} \le 1. \tag{27}$$

B.1.2 Autocovariance Function

Due to independence of $h_{1,n}, \ldots, h_{k,n}$ the following holds for $s = 1, 2, \ldots$:

$$\mathrm{cov}(h_n, h_{n-s}) \;=\; \sum_{i=1}^{k} \mathrm{cov}(h_{i,n}, h_{i,n-s}).$$

The autocovariance function of $h_{i,n}$ is given by:

$$\begin{aligned}\mathrm{cov}(h_{i,n}, h_{i,n-s}) \;&=\; \mathrm{E}(\mathrm{cov}(h_{i,n}, h_{i,n-s}|\sigma^2_{i,n-s-1})) \\ &+\; \mathrm{cov}(\mathrm{E}(h_{i,n}|\sigma^2_{i,n-s-1})\,\mathrm{E}(h_{i,n-s}|\sigma^2_{i,n-s-1})).\end{aligned} \tag{28}$$

The first term in (28) disappears, because the increases in integrated volatility $h_{i,n-s}$ and $h_{i,n}$ in the intervals $[(n-s-1)\Delta,(n-s)\Delta]$ and $[(n-1)\Delta,n\Delta]$ are conditionally independent given the actual volatility $\sigma^2_{i,n-s-1}$. From (16) we obtain:

$$\begin{aligned}\mathrm{E}(h_{i,n}|\sigma^2_{i,n-s-1}) &= \alpha_i\Delta + \mathrm{E}(\sigma^2_{i,n-1}-\alpha_i|\sigma^2_{i,n-s-1})\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i}\ ,\\ \mathrm{E}(h_{i,n-s}|\sigma^2_{i,n-s-1}) &= \alpha_i\Delta + (\sigma^2_{i,n-s-1}-\alpha_i)\frac{1-\exp(-\lambda_i\Delta)}{\lambda_i}.\end{aligned}$$

From (17) we obtain by taking iterated expectations:

$$\mathrm{E}(\sigma^2_{i,n-1}-\alpha_i|\sigma^2_{i,n-s-1}) \;=\; \exp(-\lambda_i s\Delta)(\sigma^2_{i,n-s-1}-\alpha_i).$$

Therefore:

$$\mathrm{cov}(h_{i,n},h_{i,n-s}) \;=\; \exp(-\lambda_i s\Delta)(1-\exp(-\lambda_i\Delta))^2\frac{\alpha_i\tau_i^2}{2\lambda_i^3}. \tag{29}$$

B.2 Marginal Moments, Autocorrelation and Squared Autocorrelation of the Returns

Based on representing the returns as $y_n=\sqrt{h_n}z_n$, where z_n is an iid standard normal sequence, we obtain the higher order moments of the marginal distribution of the return process y_n as:

$$\mathrm{E}(y_n^k) = \mathrm{E}(\mathrm{E}(y_n^k|h_n)) = m_k\,\mathrm{E}(h_n^{k/2}), \tag{30}$$

with m_k being the kth moment of the standard normal distribution. Furthermore for $s=1,2,\ldots$ the following holds:

$$\mathrm{E}(y_n^k y_{n-s}^l) = \mathrm{E}(\mathrm{E}(y_n^k y_{n-s}^l|h_n,h_{n-s})) = m_k m_l\,\mathrm{E}(h_n^{k/2}h_{n-s}^{l/2}). \tag{31}$$

B.2.1 The marginal moments

Since $m_1=m_3=0$, it follows immediately from (30) that

$$\mathrm{E}(y_n) = \mathrm{E}(y_n^3) = 0. \tag{32}$$

From (22) and (30) we obtain:

$$\mathrm{var}(y_n) = \mathrm{E}(y_n^2) = \mathrm{E}(h_n) = \Delta\sum_{i=1}^{k}\alpha_i. \tag{33}$$

The kurtosis is given by:

$$\mathrm{kur}(y_n) = \frac{\mathrm{E}(y_n^4)}{\mathrm{var}(y_n)^2} = \frac{3\,\mathrm{E}(h_n^2)}{\mathrm{E}(h_n)^2} = 3+3\frac{\mathrm{var}(h_n)}{\mathrm{E}(h_n)^2}. \tag{34}$$

B.2.2 Autocorrelation and Squared Autocorrelation

It follows immediately from (31) with $k = l = 1$ that y_n is uncorrelated. With $k = l = 2$ we obtain that the autocovariance function of the squared returns y_n^2 is equal to the autocovariance function of h_n:

$$\begin{aligned}\operatorname{cov}(y_n^2, y_{n-s}^2) &= \operatorname{E}(y_n^2 y_{n-s}^2) - \operatorname{E}(y_n^2)\operatorname{E}(y_{n-s}^2) = \operatorname{E}(h_n h_{n-s}) - \operatorname{E}(h_n)\operatorname{E}(h_{n-s}) \\ &= \operatorname{cov}(h_n, h_{n-s}).\end{aligned}$$

The variance of the squared returns is given by:

$$\operatorname{var}(y_n^2) = \operatorname{E}(y_n^4) - \operatorname{E}(y_n^2)^2 = 3\operatorname{E}(h_n^2) - \operatorname{E}(h_n)^2 = 3\operatorname{var}(h_n) + 2\operatorname{E}(h_n)^2.$$

Therefore the autocorrelation of the squared returns is given by:

$$\operatorname{corr}(y_n^2, y_{n-s}^2) = \frac{\operatorname{cov}(y_n^2, y_{n-s}^2)}{\operatorname{var}(y_n^2)} = \frac{\operatorname{cov}(h_n, h_{n-s})}{3\operatorname{var}(h_n) + 2\operatorname{E}(h_n)^2}. \tag{35}$$

C Optimal Level of Non-Centering

To obtain in (12) an optimal level ν of non-centering with respect to the mean we apply Papaspiliopoulos et al. (2003) to following hierarchical model which is already centered with respect to the variance:

$$\begin{aligned} y_n &= h_n\varepsilon_n = \sqrt{A^\gamma X_n}\varepsilon_n, \qquad \varepsilon_n \sim \mathcal{N}(0,1), \\ X_n &= \frac{h_n}{A^\gamma}\,. \end{aligned} \tag{36}$$

The resulting partially centered variable is given by $h_{n,PC} = X_n - \nu E(X_n)$ where $\nu = (1 + I(y_n)\sigma_x^2)^{-1}$ with $I(y_n)$ being the Fisher information and σ_x^2 being equal to $\operatorname{var}(X_n)$.

The log-likelihood for y_n in (36) is given by:

$$\ell(y_n, X_n, \theta) = -0.5\log 2\pi - 0.5\log A^\gamma X_n - 0.5\frac{y_n^2}{A^\gamma X_n}\,. \tag{37}$$

First, we derive the expected Fisher information using $\operatorname{E}(y_n^2) = A^\gamma X_n$:

$$\operatorname{E}\left(-\frac{\partial^2}{\partial X_n^2}\ell(y_n, X_n, \theta)|X_n\right) = -\frac{1}{2X_n^2} + \frac{\operatorname{E}(y_n^2)}{A^\gamma X_n^3} = \frac{1}{2X_n^2} = \frac{A^{2\gamma}}{2h_n^2}. \tag{38}$$

Since this is still a random variable, it is substituted by the expected value to obtain $I(y_n)$. By combining $h_n \approx \Delta\sigma_n^2$ with the marginal distribution $\sigma_n^2 \sim \operatorname{Gam}(\alpha b, b)$, where $b = 2\lambda/\tau^2$, we obtain that, approximately, $h_n \sim \operatorname{Gam}(\alpha b, b/\Delta)$. As long as $\tilde{a} = 2\alpha\lambda/\tau^2 > 2$, the expectation of $1/h_n^2$ exists and is given by:

$$\operatorname{E}\left(\frac{1}{h_n^2}\right) = \frac{\Gamma(\tilde{a}-2)b^2}{\Gamma(\tilde{a})\Delta^2}, \tag{39}$$

and the expectation in (38) is approximated by:

$$I(y_n) \approx \frac{A^{2\gamma}b^2}{2\Delta^2(\tilde{a}-1)(\tilde{a}-2)}. \tag{40}$$

The same approximate marginal Gamma distribution is used to approximate $\sigma_x^2 = \text{var}(X_n) = \text{var}(h_n)/A^{2\gamma}$:

$$\sigma_x^2 \approx A^{-2\gamma}\alpha\Delta^2/b. \tag{41}$$

Thus, for $\tilde{a} = 2\alpha\lambda/\tau^2 > 2$,

$$\nu = (1 + I(y_n)\sigma_x^2)^{-1} = \frac{2(\tilde{a}-1)(\tilde{a}-2)}{2(\tilde{a}-1)(\tilde{a}-2)+\tilde{a}} \ . \tag{42}$$

If in (36) h_n is substituted by $\sigma_n^2\Delta$ we obtain following hierarchical model:

$$\begin{aligned} y_n &= \sqrt{B^\gamma X_n/\Delta}\varepsilon_n, \qquad \varepsilon_n \sim \mathcal{N}(0,1), \\ X_n &= \frac{\sigma_n^2\Delta}{B^\gamma} \ . \end{aligned}$$

The resulting partially centered variable is given by $\sigma^2_{n,PC} = X_n - \nu E(X_n)$ where ν is the same as in (42).

D Density Approximation with the Multivariate Normal Distribution

This section investigates the effect of approximating the transition density $p(h_{i,n}, \sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i)$ by means of a multivariate normal distribution. Thus we compare results obtained from a numerical inversion of $p(h_{i,n}, \sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i)$ by means of

$$p(h_{i,n}, \sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) \approx \frac{1}{(2\pi)^2}\sum_u\sum_v \exp(-\imath(uh_{i,n} + v\sigma^2_{i,n}))\phi(u, v|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i)\ \Delta_u\Delta_v$$

against a bivariate normal approximation $p(h_{i,n}, \sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i) \approx \mathcal{N}(\mu_{i,n}, \Sigma_{i,n})$. The components of $\mu_{i,n}$ are given by (16) and (17), respectively. The elements of the variance-covariance matrix $\Sigma_{i,n}$ follow from (19), (20) and (21).

For the numerical inversion a grid, symmetric with respect to the origin, of $(M+1)\times(M+1)$ pairs (u, v) is used and the step-widths Δ_u and Δ_v are set to one. Using $M = 100$ resulted in sufficiently good inversion results. We compare the normal approximation against this benchmark. Compared to the numerical inversion, the numerical effort of the normal approximation is approximately $1/M^2$ times smaller.

First, we compare the individual transition densities the log scale, i.e. $\log p(h_{i,n}, \sigma^2_{i,n}|\sigma^2_{i,n-1}, \boldsymbol{\theta}_i)$. We generate paths with the model parameters of Section 4.1 with 500 observations. For the individual densities the following differences between the numerical inversion and the normal approximation are observed: a mean difference of 0.0154, a median difference of 0.0027, and 5%, 10%, 90% and 95% quantiles of -0.4053, -0.2684, 0.3332 and 0.4705, respectively. Since the log-density can be close to zero and the ratio of the transition densities is important in the Metropolis Hastings algorithm, the results in percentage terms can show a distorted picture of the approximation problem. Therefore, we also calculated the deviations of the individual transition densities in percentage terms. Here we observe a mean deviation of

-0.0285%, and a median of -0.0015%. While the 90% quantile of the difference in the densities is 0.2687%, the 5% and the 95% quantiles are -0.5469% and 0.3499%, i.e. there can be substantial differences. These differences are of course caused by approximation errors in the tails of the normal approximation, but also by numerical difficulties for low probabilities with the inversion. This second aspect is also essential when running an MCMC sampler. Problems with inversion integrals for small probabilities can dominate the updating properties of the MCMC sampler.

In addition we also compare the differences in the total likelihood,

$$\prod_{n=1}^{N} p(h_{i,n}, \sigma_{i,n}^2 | \sigma_{i,n-1}^2, \boldsymbol{\theta}_i)$$

for all $N = 500$ observations, because this determines the error in the acceptance probability of the Metropolis-Hastings algorithm. It turned out that the total likelihood differed by about 3%. From this analysis we conclude that a density approximation by a bivariate normal distribution is sufficient.

Remark Lamoureux and Paseka (2005) derive the Fourier transform of the joint density of yields and volatility. By means of analytically inverting the Fourier transform they can reduce the dimension of the inversion integral to a one-dimensional inversion integral. Here the Fourier transform exhibits a different structure, such that the calculations of Lamoureux and Paseka (2005) cannot be applied.